Systematic review and meta-analysis

David R Jones

Centre for Biostatistics and Genetic Epidemiology, University of Leicester, UK

Aims of the session

- To indicate the rationale for systematic review and meta-analysis
- To outline the main steps in systematic reviewing
- To outline steps in and provide initial experience in performing a simple meta-analysis, in a worked example from environmental epidemiology

Levels of evidence

see Centre for Evidence-Based Medicine(CEBM) website [1]

Matching the source, type and quality of evidence to the question to be answered is important, but as will be clear from the CEBM website, systematic reviews are almost always seen as the top level evidence.

Distinguishing systematic review (SR) and meta-analysis (MA)

The terms SR and MA do not have universal definitions, but here:

Distinguishing systematic review (SR) and meta-analysis (MA)

The terms SR and MA do not have universal definitions, but here:

- SR: Uses explicit, transparent, repeatable criteria in identifying all evidence relevant to clearly focused question
  - Explicit about:
    - types of study
    - participants
    - interventions
    - outcome measures included

Distinguishing systematic review (SR) and meta-analysis (MA)

The terms SR and MA do not have universal definitions, but here:

- SR: Uses explicit, transparent, repeatable criteria in identifying all evidence relevant to clearly focused question
  - Explicit about:
    - types of study
    - participants
    - interventions
    - outcome measures included
- MA: *quantitative* combination of results of primary studies
  - typically a weighted average
  - exploration and modelling of heterogeneity,... and bias?

Meta-analysis

- Derivation:
  - $\mu\epsilon\tau\alpha$, meaning 'after', 'above', 'transcending'

- Definition:

  The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the finding.

  Glass, 1976

Bias in meta-analysis: a recent view

.. meta-analysis (a procedure in which, metaphorically speaking, apples, oranges and an unspecified fruit are confidently blended on the assumption that the resulting liquid has revelatory properties).

David Smail (Times Higher Education, 12 Nov 09, p46)

Why should we do SRs and MAs?

- Systematic review:
  - To collate a *complete* set of relevant study results
  - To make review process and criteria
    - explicit
    - transparent
    - repeatable
    - updatable
- Meta-analysis:
  - To *quantify* effect sizes and their uncertainty
  - To reduce sampling variations
  - To facilitate synthesis and interpretation of several/many study results

Good SR and MA practice[1]

1. Specify scope of SR/MA, target question and method in a protocol
2. Compile a complete set of relevant studies using electronic databases, reference explosion,..
3. Define common, compatible outcome and exposure measures, and covariates
4. Extract standard data items, in blinded and duplicated way
5. If appropriate, meta-analyse allowing for sources of variation and checking for bias
6. Perform sensitivity analyses to study quality, and potential biases

[1] see Sutton [2], MOOSE guidelines [3], Nicholson [4]

Meta-analysis: methods and approaches

- Vote counting
- Combining p values
- **Combining estimates of effect sizes:**
  - **fixed effect model**
  - **random effects model**
- Bayesian versions/extensions of 3 above

Meta-analysis: methods and approaches

- Vote counting
- Combining p values
- **Combining estimates of effect sizes:**
  - **fixed effect model**
  - **random effects model**
- Bayesian versions/extensions of 3 above
- Pooling of individual subject data [IPD] cf. summary study level data as above

Worked MA example: Melanoma and sun exposure

- **Disclaimer:** This is a simple illustrative, example, based on a subset of case-control studies reviewed in an earlier meta-analysis (Elwood and Jopson 1997[5]). It is NOT a definitive review or meta-analysis of the topic.

- Here consider **intermittent** exposure to sunlight, and compare just the highest vs. lowest reported exposure subgroups in each paper.
- Need to extract 4 numbers from each paper to allow the odds ratio and its standard error to be calculated:

	Melanoma (cases)	**No melanoma (controls)**
Highest exposure	a	b
Lowest exposure	c	d

Standard errors of and confidence intervals for odds ratios

If we observe data:

	Disease cases	**No disease controls**
Exposed	a	b
Not exposed	c	d

Estimate odds ratio by $\frac{ad}{bc}$

Then log(OR) has approximate variance (Woolf's method):

$$\left\{\frac{1}{a}\right\} + \left\{\frac{1}{b}\right\} + \left\{\frac{1}{c}\right\} + \left\{\frac{1}{d}\right\}$$

Studies of melanoma and intermittent sun exposure

from Elwood and Jopson, 1997[5]

Author	**a**	**b**	**c**	**d**
Grob	46	11	87	199
Dubin	86	93	103	174
Elwood	77	59	172	230
Osterlind	432	791	42	135
Zaridize	49	35	7	14
Rodenas	38	17	56	107
Green	16	6	18	13
Dubin '86	372	110	397	110
Aubier	14	4	59	90

Summarising results of each primary study (as OR and s.e.)

For the 1st study (Grob):

	Disease cases	**No disease controls**
Exposed	46	11
Not exposed	87	199

Estimate odds ratio by $\frac{46x199}{87x11} = 9.57$

On log scale $variance[log(OR)] \approx \left\{\frac{1}{46}\right\} + \left\{\frac{1}{11}\right\} + \left\{\frac{1}{87}\right\} + \left\{\frac{1}{199}\right\} = 0.292$

Then $s.e.[log(OR)] = 0.359$ and 95% confidence interval for log (OR) easily calculated.

Summarising results of each primary study (as OR and s.e.)

For the 1st study (Grob):

	Disease cases	**No disease controls**
Exposed	46	11
Not exposed	87	199

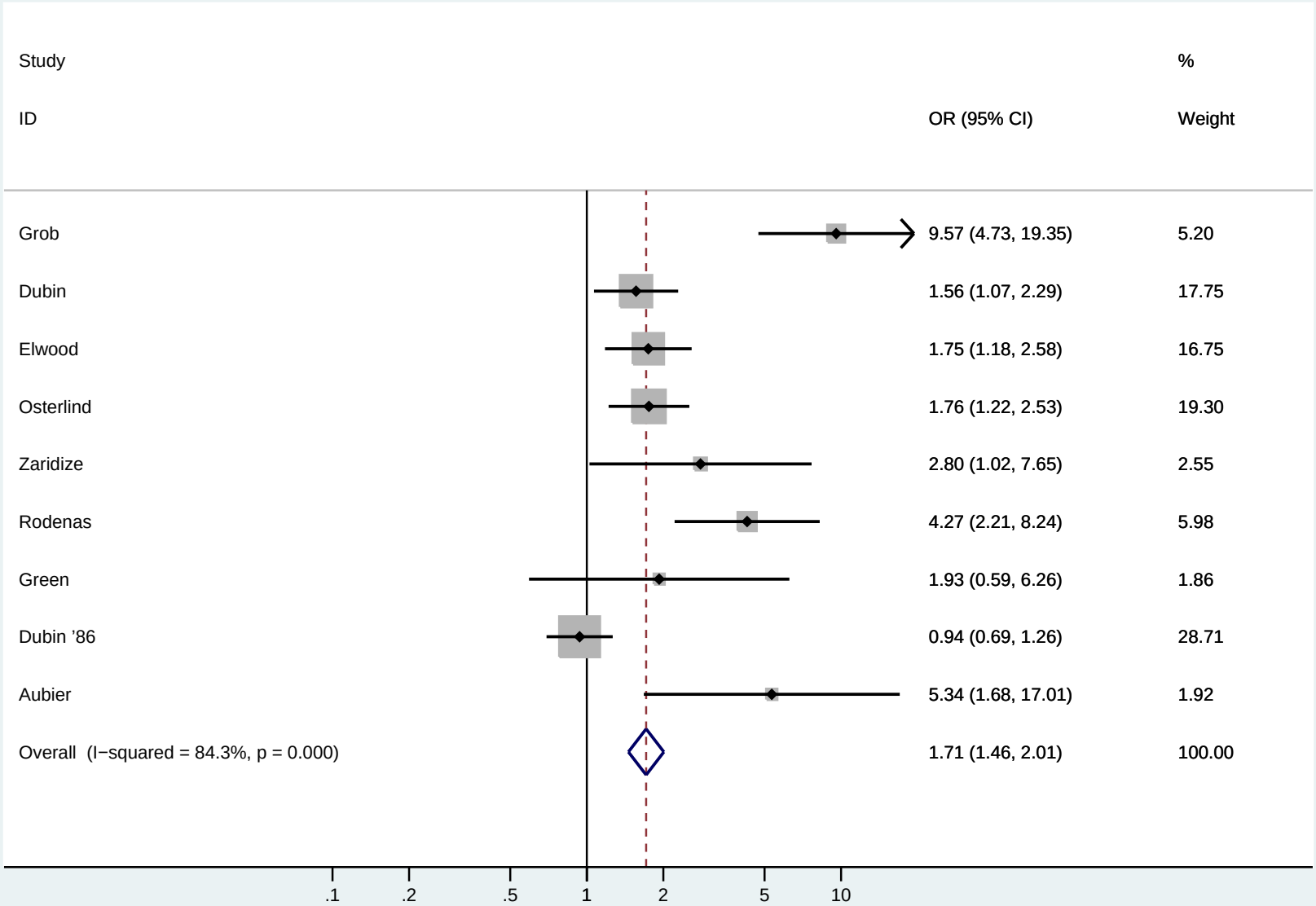
Estimate odds ratio by $\frac{46x199}{87x11} = 9.57$

On log scale $variance[log(OR)] \approx \left\{\frac{1}{46}\right\} + \left\{\frac{1}{11}\right\} + \left\{\frac{1}{87}\right\} + \left\{\frac{1}{199}\right\} = 0.292$

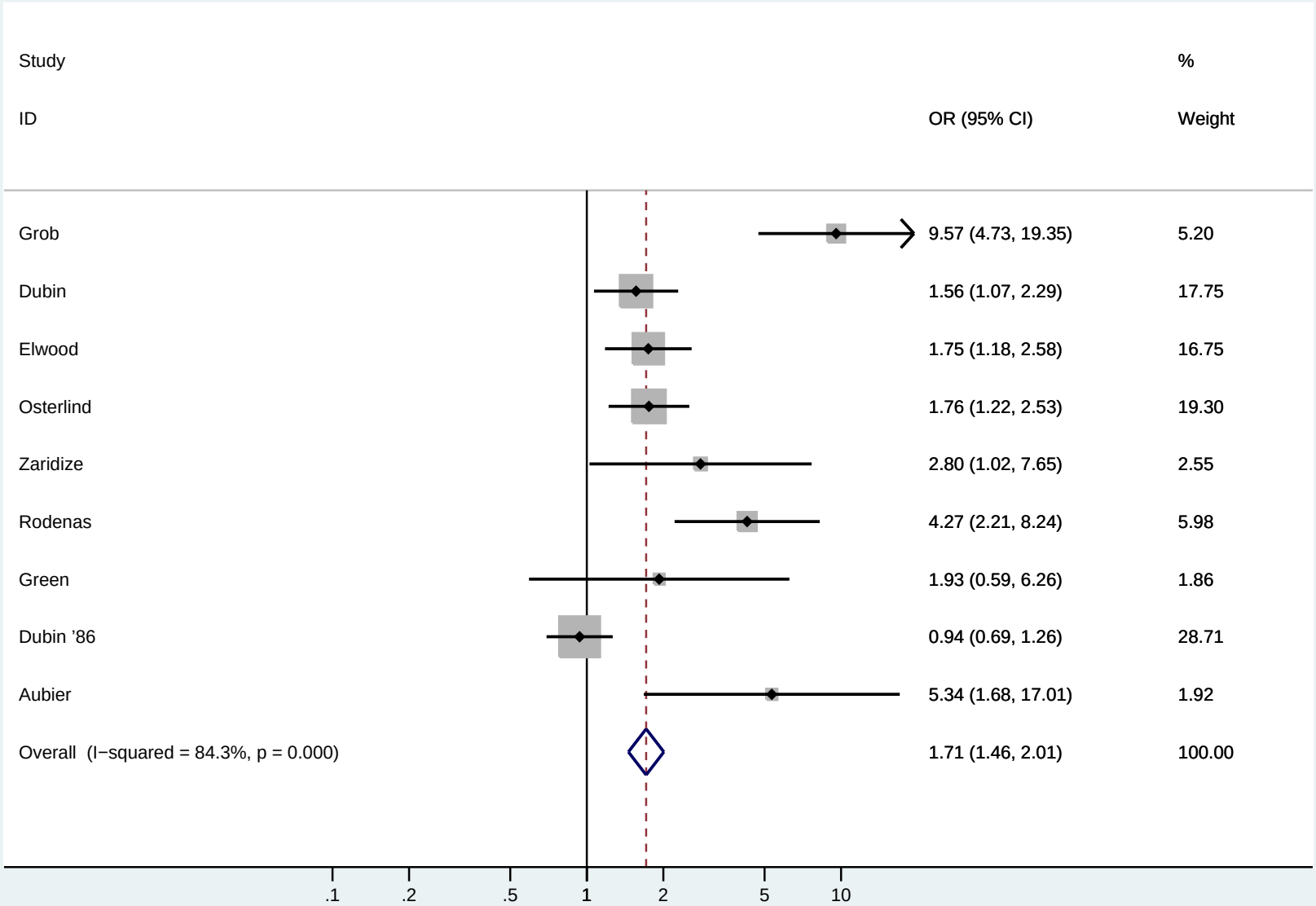
Then $s.e.[log(OR)] = 0.359$ and 95% confidence interval for log (OR) easily calculated.

- In practice, use package such as Stata, Revman, R, MIX 2.0 to do calculations (sometimes using more exact formulae) and display results.

Forest plot for melanoma example



Forest plot for melanoma example, *using fixed effect model*



Fixed effect meta-analysis models

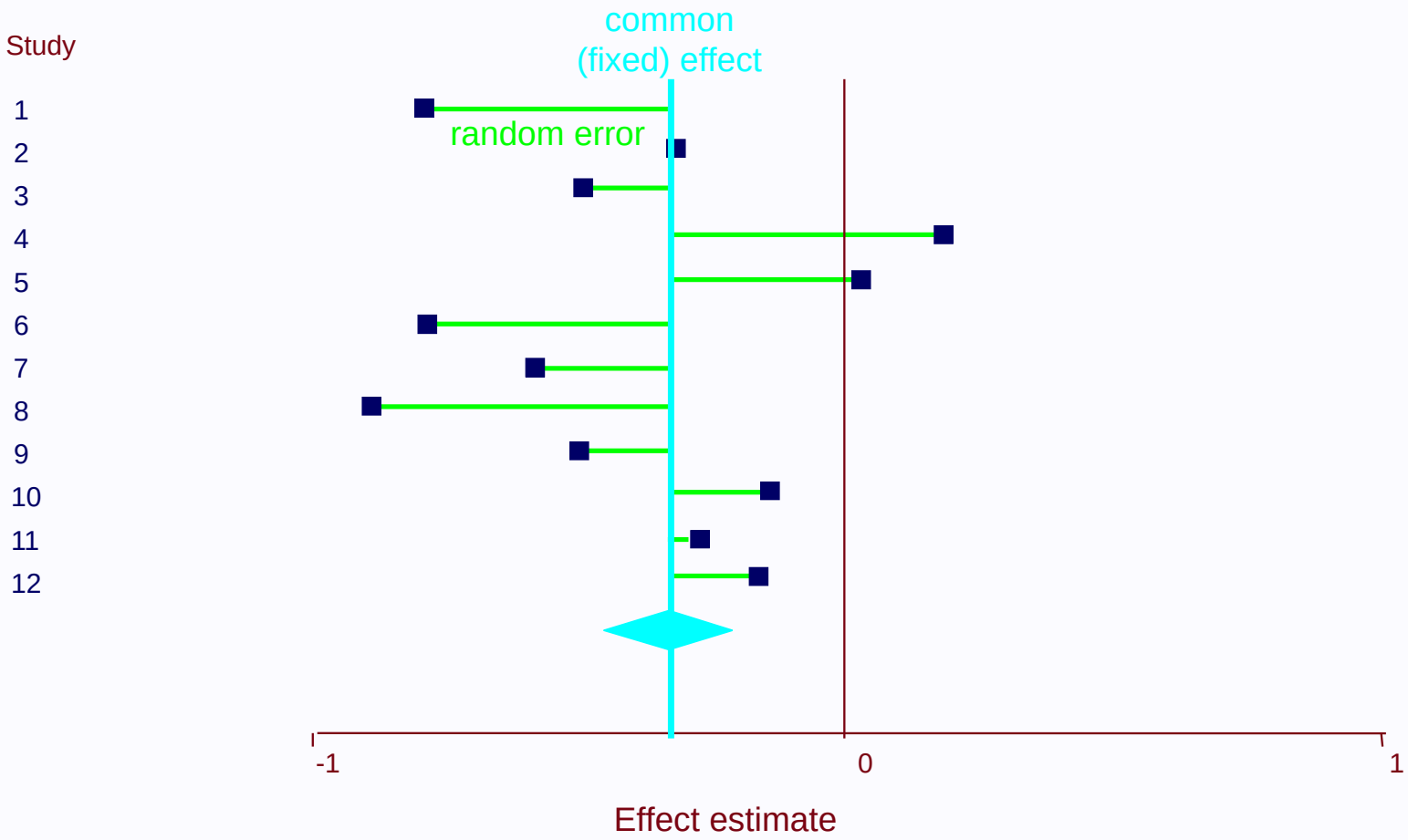
- Assumes **true** effect is the **same** in all primary studies

Fixed effect meta-analysis models

- Assumes **true** effect is the **same** in all primary studies
- Observed value in study i = true value + error in study i
- $Y_i = \theta + e_i$ *with* $\theta_i = \theta$ *for all* i
- Estimate a common effect (outcome value) with a **weighted average** of the results from primary studies
- Weight value from ith study with its precision ($w_i = \frac{1}{variance(\hat{\theta}_i)}$)
- *Common value* $\hat{\theta} = \frac{\Sigma w_i \hat{\theta}_i}{\Sigma w_i}$
- Hence results from large studies have small variances and hence are given heavy weights, and so on.

How does a fixed effect model work?

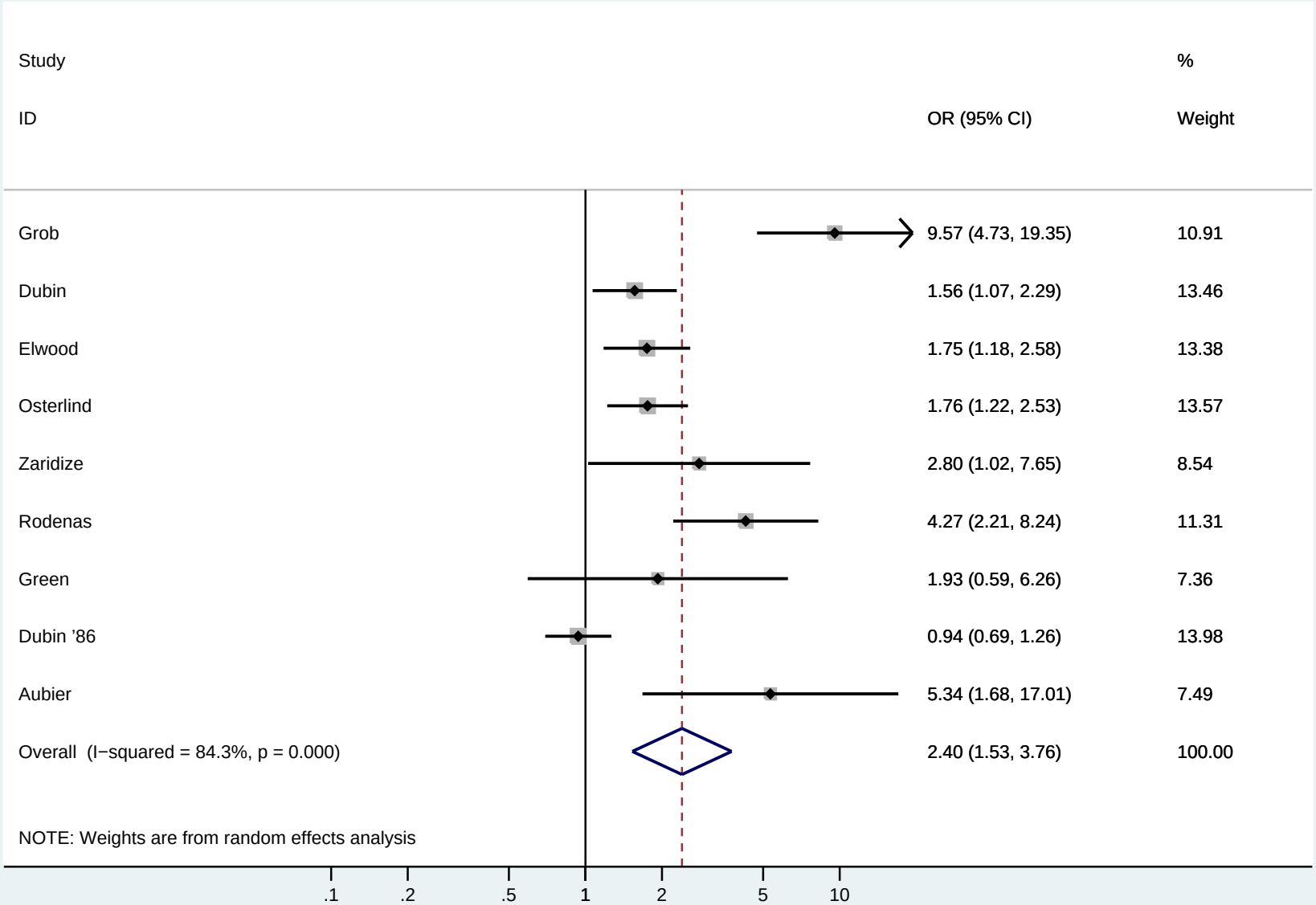
with acknowledgement to Julian Higgins



Heterogeneity

- Fixed effect model assumes all studies are estimating the same true effect size
- This is usually *not* a reasonable assumption, since primary studies may differ in design, conduct, context,...
- Can do a statistical test to see if observed heterogeneity is plausible by chance alone...
- ...but the the test itself has low power (so may not always detect heterogeneity when present)
- ... and it is more valuable to **account for and explore** sources of the heterogeneity

Melanoma example: *Random* effects model



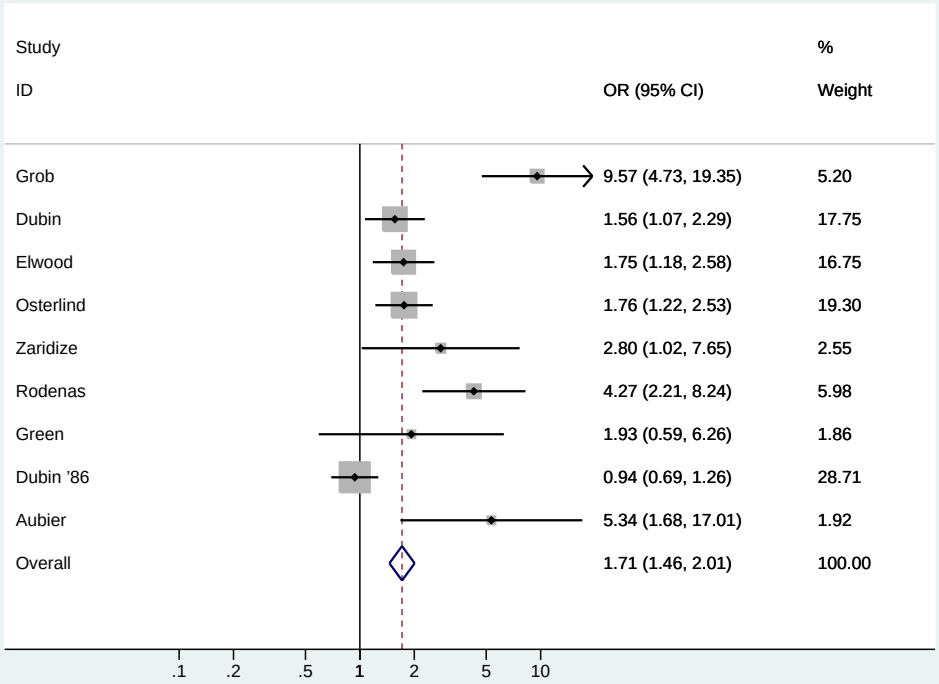
Random effects MA models

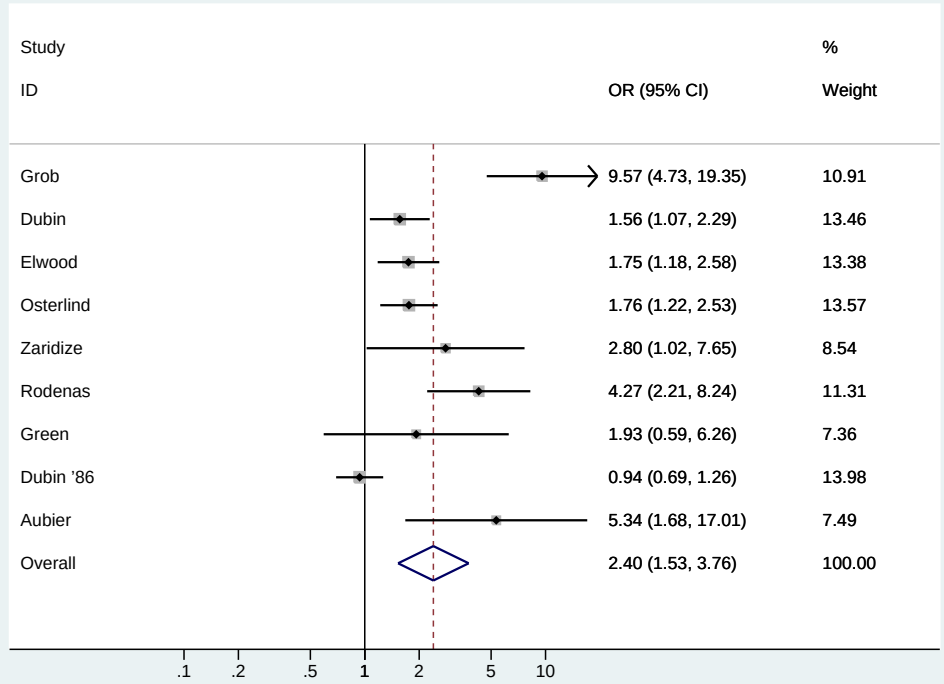
- Does not assumes **true** effect is the **same** in all primary studies,
- Instead assumes true value varies between studies, but according to a pre-specified distribution
- $Y_i = \theta_i + e_i$ *with* $\theta_i \sim N(\theta, \tau^2)$ *for all* i
- Estimate a common effect (outcome value) $[\theta]$ as the mean of the assumed distribution with a **weighted average** of the results from primary studies
- *Weight value from ith study with a combination of its precision and an estimate of the between study variability* ($w_i = \frac{1}{var(\hat{\theta}_i)+\tau^2}$)

Random vs. fixed effects models

- Random effects point estimate are often (but not always) similar to the fixed effect estimate
- Confidence intervals from random effects are usually *wider* than those from fixed effects
- Primary studies are given **more equal weighting** in random effects than in fixed effect MAs
- Debate over use of fixed or random effects continues [6]..

Melanoma example: fixed vs. random effects

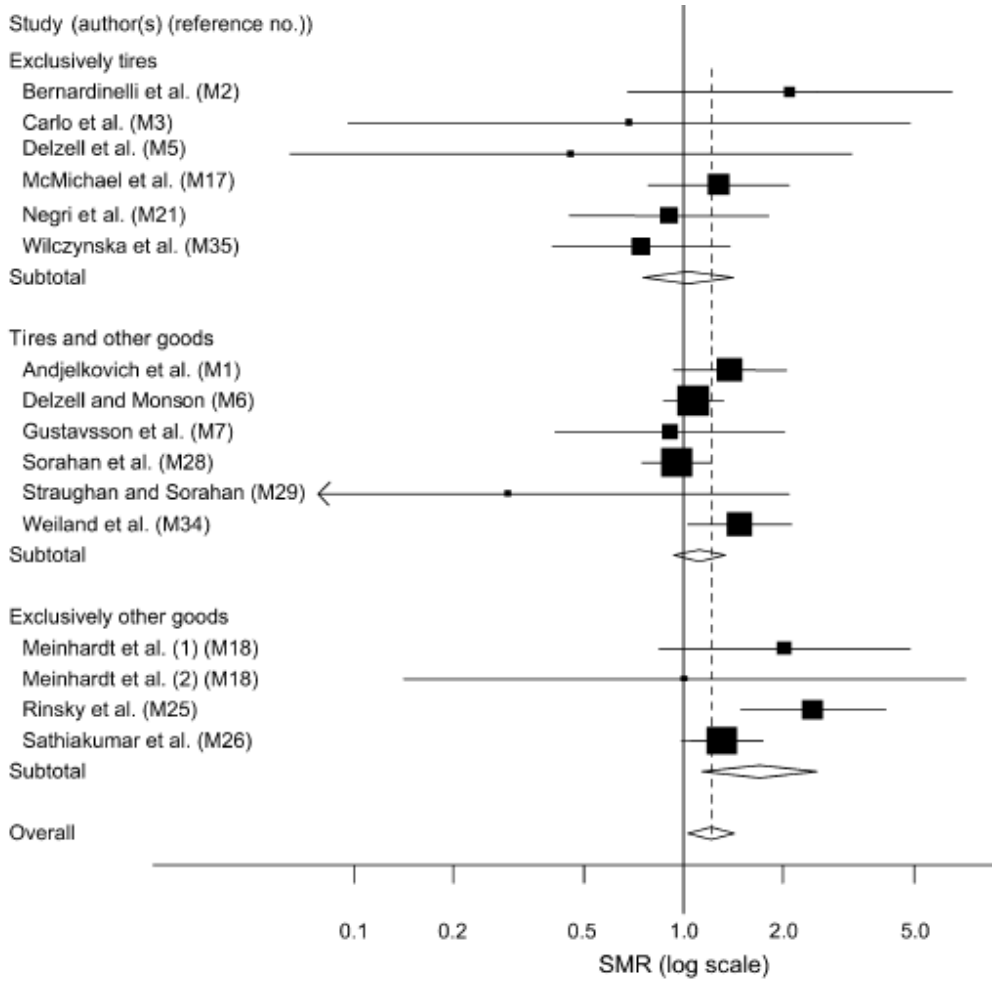




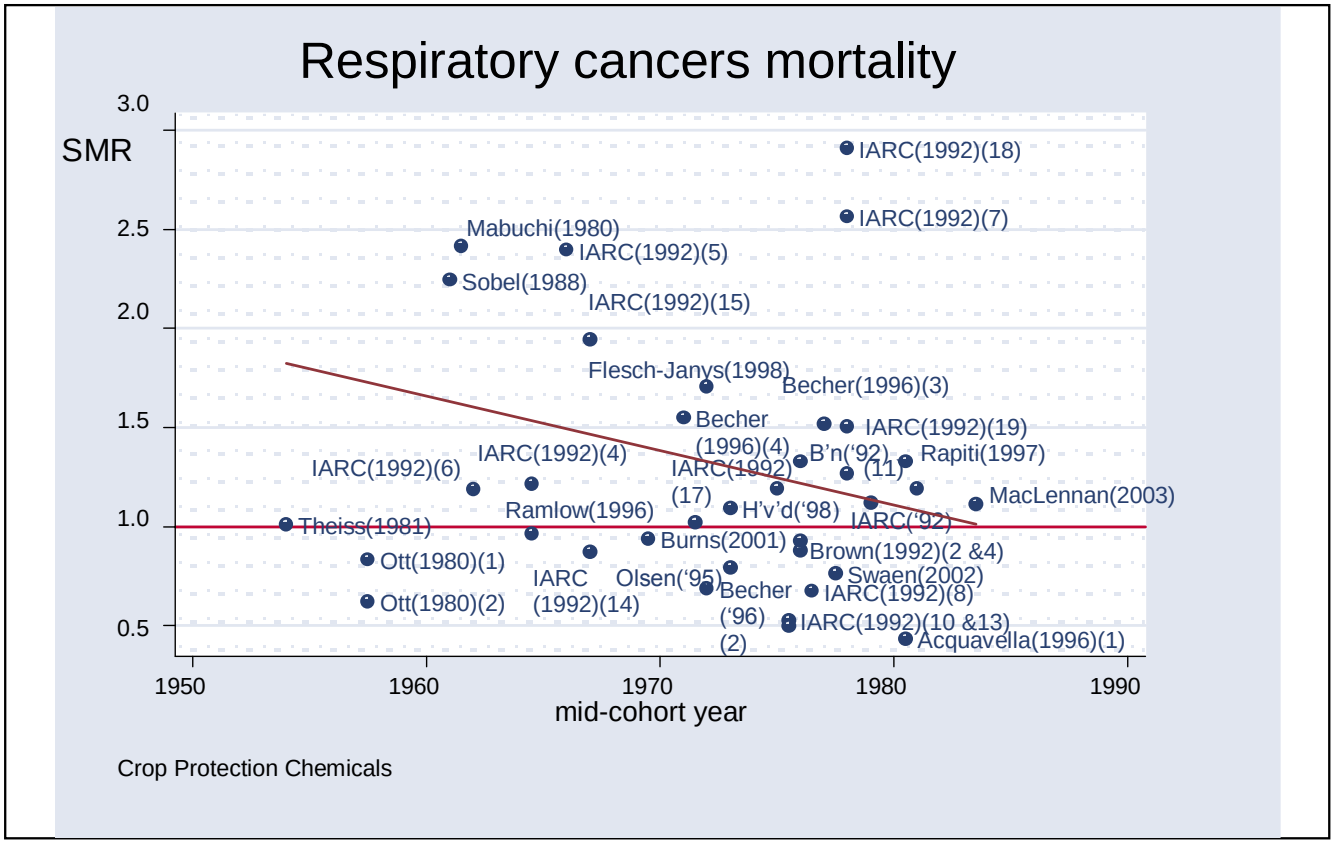
Exploring heterogeneity

- Random effects models can **account for** heterogeneity but do not **explain** it
- Subgroup analyses and/or meta-regression can help identify sources of heterogeneity and explain it
- Subgroup analyses may repeat the MA in
  - subsets of the studies which have particular characteristics (e.g. UK studies, studies since 2005, studies with follow up of more than 10 years, etc), or
  - subsets of subjects within the studies (e.g. subjects aged 50+,less than 50, etc), perhaps to study effect modification
- *Meta-regression explores and adjusts for differences between studies by including covariates characterising the studies as above*

Subgroup analysis example: leukemia mortality by industry sub-type (Alder et al, 2006[7])



Meta-regression example: Respiratory cancer mortality by year of publication (Jones et al, 2009[8])



Meta-analysis: problems

- Publication and related biases: selective reporting of primary studies, or specific results therein, dependant on significance or 'interestingness' of results
- Primary studies vary in respect of
  - quality:

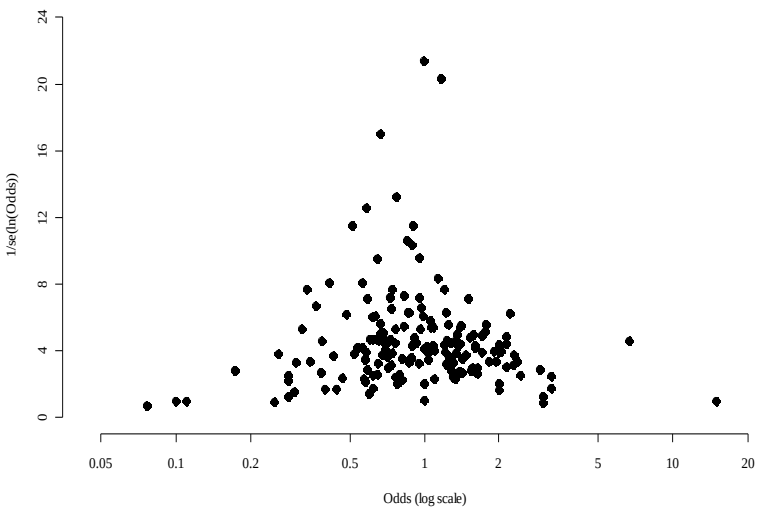
    ..apples and oranges and the occasional lemon..

  - context, subject/population characteristics, design - which *may* militate against combining results from them

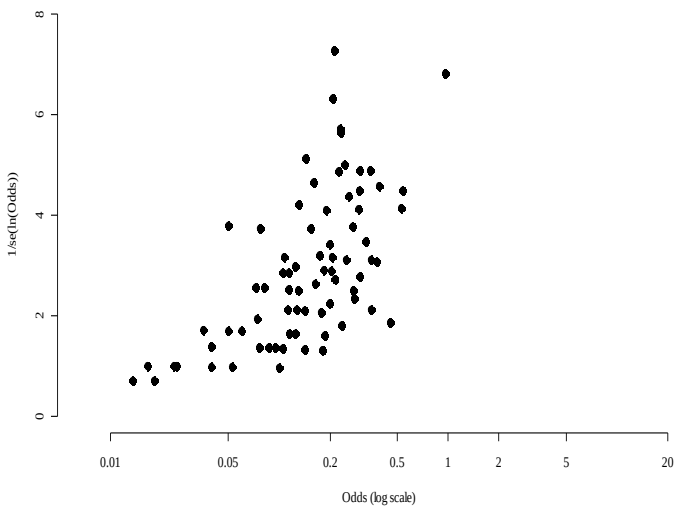
Funnel plots (to investigate publication bias)

- Plot of some measure of study size v. outcome:

No evidence of bias

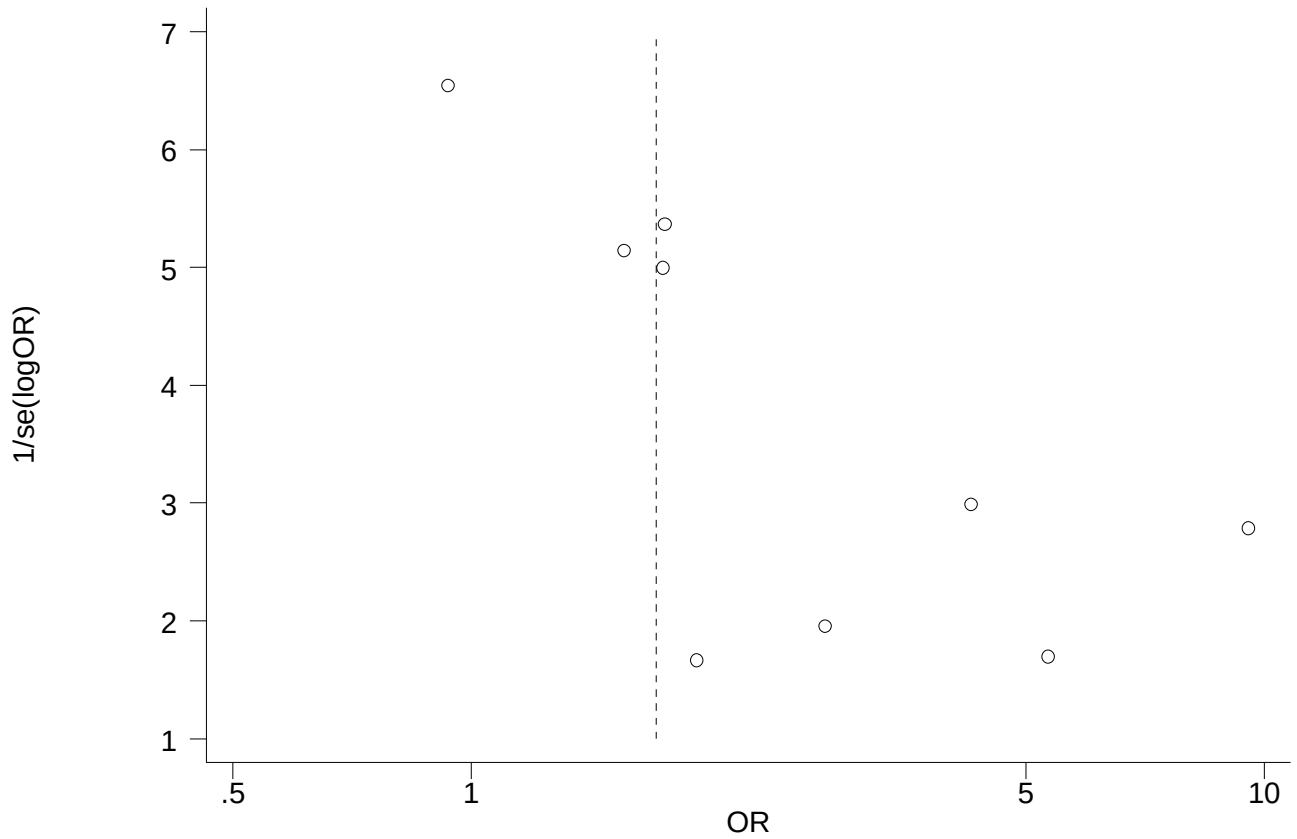


Evidence of bias



Epidemiology & Public Health, Leicester University & Nuffield College, Oxford

Funnel plot: melanoma example



On data quality

Public agencies are very keen on amassing statistics - they collect them, add them, raise them to the n'th power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of those figures comes in the first instance from the village watchman, who just puts down what he damn well pleases.

Sir Josiah Stamp (1880-1941)

or:

Garbage in, garbage out

Dealing with study quality in MAs

- Many checklists and scales for evaluation of quality of epidemiological studies (and other study types) exist
- Substantial evidence that relationships between quality scores and study results vary (so no easy route to allowing for study quality)
- Sensitivity of results to exclusion of low quality studies should, however, be explored
- Methods for adjustment of primary study results for (internal and external) bias before synthesis are current research topics [9]

Sensitivity analyses in MAs

- provide an approach to checking how robust results/conclusions of an MA are to assumptions made therein.
- are strongly recommended.
- allow investigation of influence of:
  - study quality
  - inclusion/exclusion criteria
  - uncertainties in data extraction
  - missing values (due to publication bias or other causes)
  - ongoing studies
  - ... and so on

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes
Expert judgement of weight of evidence for qualitative evaluation	Yes	No

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes
Expert judgement of weight of evidence for qualitative evaluation	Yes	No
Quantitative synthesis of primary study results	No	Yes

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes
Expert judgement of weight of evidence for qualitative evaluation	Yes	No
Quantitative synthesis of primary study results	No	Yes
Narrative or qualitative summary and evaluation	Yes	Yes

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes
Expert judgement of weight of evidence for qualitative evaluation	Yes	No
Quantitative synthesis of primary study results	No	Yes
Narrative or qualitative summary and evaluation	Yes	Yes
Sensitivity analyses	No	Yes

Narrative vs. systematic review

	Narrative review	**SR and/or MA**
Transparent compilation of comprehensive set of primary study data	?/No	Yes
Common set of outcomes etc identified	?	Yes
Expert judgement of weight of evidence for qualitative evaluation	Yes	No
Quantitative synthesis of primary study results	No	Yes
Narrative or qualitative summary and evaluation	Yes	Yes
Sensitivity analyses	No	Yes
Explicit reporting of review methodology	?	Yes

Summary

- Systematic review helps avoid evidence and reviewer biases by using explicit, transparent and repeatable criteria in identifying all evidence relevant to clearly focused and specified questions
- Meta-analysis allows quantification of effects with more precision than from individual studies, and better exploration of heterogeneity and biases

Summary

- Systematic review helps avoid evidence and reviewer biases by using explicit, transparent and repeatable criteria in identifying all evidence relevant to clearly focused and specified questions

- Meta-analysis allows quantification of effects with more precision than from individual studies, and better exploration of heterogeneity and biases, but *beware its uncritical use*

References

[1] Oxford Centre for Evidence-Based Medicine. Steps in finding evidence ("levels") for different types of questions. *See: http://www.cebm.net/index.aspx?o=5653*, accessed 27Aug2010.

[2] A. J. Sutton, K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. *Methods for meta-analysis in medical research*. London, John Wiley, 2000.

[3] D. F. Stroup, J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, D. Rennie, D. Moher, B. J. Becker, T. A. Sipe, and S. B. Thacker. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*, 283(15):2008, 2000. [MOOSE].

[4] PJ Nicholson. How to undertake a systematic review in an occupational setting. *Occupational and Environmental Medicine*, 64(5):353, 2007.

[5] J. M. Elwood and J. Jopson. Melanoma and sun exposure: an overview of published studies. *International Journal of Cancer*, 73(2):198–203, 1997.

[6] J. P. T. Higgins, S. G. Thompson, and D. J. Spiegelhalter. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159, 2009.

[7] N. Alder, J. Fenty, F. Warren, AJ Sutton, L. Rushton, DR Jones, and KR Abrams. Meta-analysis of mortality and cancer incidence among workers in the synthetic rubber-producing industry. *American Journal of Epidemiology*, 164(5):405, 2006.

[8] DR Jones, AJ Sutton, KR Abrams, J. Fenty, F. Warren, and L. Rushton. Systematic review and meta-analysis of mortality in crop protection product manufacturing workers. *Occupational and Environmental Medicine*, 66(1):7, 2009.

[9] R.M Turner, D.J. Spiegelhalter, G.C.S. Smith, and S.G. Thompson. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 172(1):21–47, 2009.

Contact details

Comments and queries to

drj@le.ac.uk